

# Trợ lý ảo cho Robot tiếp tân Khách sạn Dựa trên Lý thuyết sRAM và nền tảng Dialogflow

Đình Bảo Minh, Dương Văn Minh, Nguyễn Quốc Bảo, Ngô Sỹ Trung, Nguyễn Trường Sơn

Khoa Điện tử - Viễn thông

Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Email: [minhdinh@vnu.edu.vn](mailto:minhdinh@vnu.edu.vn), [meiyou712@gmail.com](mailto:meiyou712@gmail.com), [qbao1607@gmail.com](mailto:qbao1607@gmail.com), [trungns.bn@gmail.com](mailto:trungns.bn@gmail.com), [truongsonlifezero@gmail.com](mailto:truongsonlifezero@gmail.com)

**Abstract**— Robot ngày càng trở nên phổ biến trong ngành dịch vụ và du lịch. Đặc biệt, trong bối cảnh của dịch COVID-19, robot tiếp tân đã trở thành giải pháp giúp hạn chế tiếp xúc và giảm nguy cơ lây nhiễm. Là dịch vụ tư vấn và hướng dẫn khách hàng, công việc tiếp tân đòi hỏi robot phải có khả năng tương tác linh hoạt với khách hàng. Bài báo trình bày một trợ lý ảo hỗ trợ robot tiếp tân khách sạn tương tác với khách hàng dựa trên cơ sở lý thuyết “Service Robot Acceptance Model” (sRAM). Trợ lý ảo được xây dựng trên nền tảng Dialogflow của Google. Kết quả kiểm thử với nhiều kịch bản hội thoại cho thấy hệ thống có độ chính xác cao, có độ trễ thấp và có thể áp dụng trong thực tế. Một phần kết quả thực nghiệm được ghi hình và trình bày tại: [Kết quả thực nghiệm](#)

**Keywords**- Trợ lý ảo, sRAM, Dialogflow, HRI

## I. GIỚI THIỆU

Trong bối cảnh của dịch COVID-19, sự phục vụ của robot trong các lĩnh vực đời sống trở thành một trong những giải pháp giúp hạn chế tiếp xúc và giảm nguy cơ lây nhiễm dịch bệnh. Đặc biệt, với các công việc luôn phải tiếp xúc với nhiều lượt khách hàng mỗi ngày như tiếp tân khách sạn, việc thay thế con người bằng robot là rất cần thiết.

Bên cạnh đó, trong lĩnh vực khách sạn, yếu tố trải nghiệm khách hàng luôn được đặc biệt quan tâm. Việc sử dụng robot trong các dịch vụ khách sạn sẽ tạo ra trải nghiệm mới mẻ và gia tăng mong muốn đặt phòng với khách hàng mới [1].

Đối với công việc tiếp tân, việc tương tác linh hoạt với khách hàng đóng vai trò quan trọng tạo nên ấn tượng và trải nghiệm tích cực với khách hàng [2]. Điều này đòi hỏi trong quá trình tương tác với khách hàng robot tiếp tân cần tạo ra các cuộc hội thoại uyển chuyển. Các cuộc hội thoại không chỉ giới hạn ở mức đáp ứng các yêu cầu cơ bản của khách hàng mà cần phải có yếu tố cảm xúc và thậm chí có thể đưa ra gợi ý dịch vụ cho khách hàng.

Trong lĩnh vực tương tác giữa người và robot, qua tìm hiểu, nghiên cứu nhận thấy hai vấn đề sau:

**Thứ nhất:** Phần lớn các nghiên cứu về robot dịch vụ chủ yếu đề cập đến khía cạnh cơ học (thiết kế, di chuyển, định vị, tự hành,...) và nhận dạng (khuôn mặt, giọng nói, vật thể,...) [3-5]. Một số nghiên cứu chỉ xây dựng mô hình tương tác ở mức cơ bản [6, 7]. Vì vậy cần có giải

pháp để tương tác giữa người và robot trở nên linh hoạt và uyển chuyển hơn.

**Thứ hai:** Hiện nay có rất nhiều cơ sở lý thuyết nghiên cứu về lĩnh vực tương tác giữa người và robot dịch vụ. Trong đó, nổi bật hơn cả là cơ sở lý thuyết “Service Robot Acceptance Model” (sRAM) [8]. Trong một số lĩnh vực như hàng không, nhà hàng... sRAM đã được đưa vào áp dụng thực tế [9]. Tuy nhiên, đối với lĩnh vực khách sạn, sRAM chưa được sử dụng nhiều.

Với hai vấn đề nêu trên, bài báo đề xuất một trợ lý ảo hỗ trợ robot tiếp tân khách sạn dựa trên cơ sở lý thuyết sRAM. Trợ lý ảo được xây dựng thông qua việc sử dụng kết hợp các API có sẵn của Google Cloud [10] và MySQL [11]. Đồng thời thiết kế lưu đồ hội thoại riêng cho tiếp tân dựa trên nền tảng Dialogflow của Google [12]. Phần mềm mô phỏng Gazebo [13] được sử dụng để kiểm chứng khả năng tương tác của trợ lý ảo với các nền tảng robot.

Kết quả thực nghiệm cho thấy giải pháp cung cấp cho độ trễ thấp, độ chính xác cao và khả năng đáp ứng với nhiều tình huống trong hội thoại. Nhờ việc áp dụng cơ sở lý thuyết sRAM, giải pháp khắc phục được một số tồn đọng phổ biến trong robot dịch vụ.

Phần còn lại của bài báo được trình bày như sau: Phần II trình bày cơ sở lý thuyết và hướng tiếp cận. Tiếp đó, phần III mô tả thiết kế hệ thống của trợ lý ảo. Hiệu năng của trợ lý ảo được đánh giá ở phần IV. Cuối cùng, phần V trình bày kết luận và hướng phát triển.

## II. CƠ SỞ LÝ THUYẾT VÀ HƯỚNG TIẾP CẬN

### A. Cơ sở lý thuyết

Tương tác giữa người và robot dịch vụ đòi hỏi robot phải hiểu mục đích, ý nghĩa của hội thoại trong giao tiếp tự nhiên. Từ mô hình cổ điển của Davis [14] nêu ra mối liên quan giữa tính hiệu quả của robot và khả năng tiếp cận công nghệ của khách hàng, các nghiên cứu sau này đã tìm các mối liên khác bằng nhiều thực nghiệm như: tìm ra tác động tâm lý học [15], quan sát phản ứng của khách hàng [16]. Nổi bật trong số đó là mô hình “Service Robot Acceptance Model” (sRAM) [8] chỉ ra độ chấp thuận của khách hàng với robot dịch vụ dựa trên 3 yếu tố: chức năng cung cấp, cảm xúc xã hội, và

yếu tố liên quan. Nghiên cứu [17] áp dụng mô hình sRAM vào lĩnh vực khách sạn qua việc thu thập và phân tích dữ liệu phản hồi khách hàng. Qua đó đưa ra nhận xét rằng khách hàng quan tâm chủ yếu đến chức năng cung cấp, tiếp đến là cảm xúc xã hội và cuối cùng và yếu tố liên quan. Đồng thời, nghiên cứu [17] cũng đưa ra mô hình lý thuyết phân bố các thành phần cần có trong ba yếu tố trên, giúp hình dung rõ hơn về tương tác giữa người và robot trong các dịch vụ khách sạn. Cụ thể như sau:

- Yếu tố chức năng cung cấp là tổng hợp của độ thân thiện, hữu dụng mà công nghệ mang lại. Trong lĩnh vực khách sạn, yếu tố chức năng bao gồm 3 thành phần: *tiếp tân*, *dịch vụ phòng* và *hành lý*. *Tiếp tân* dùng để chỉ sự tương tác giữa người và robot trong quá trình check-in và quản lý hành lý khách hàng. Ở yếu tố *tiếp tân*, khách hàng có cảm giác rằng khả năng tương tác của robot tuy đạt mục tiêu giao tiếp nhưng còn đơn giản, chưa đáp ứng được nhiều hình thức trò chuyện. Khách hàng cũng quan tâm đến độ hiệu quả của robot hơn là vẻ bề ngoài. *Dịch vụ phòng* dùng để chỉ độ hữu ích của robot trong các công việc thường ngày của khách sạn. Ở vấn đề này, khách hàng có xu hướng tò mò và trò chuyện hơn là đưa yêu cầu cho robot. Tuy vậy, robot chưa đáp ứng được các hội thoại chưa được chuẩn bị trước, thường gặp lỗi cũng như độ trễ cao. *Hành lý* chỉ khả năng vận chuyển và lưu trữ thực hiện bởi robot. Khách hàng cảm thấy robot giúp đỡ các công việc này rất tiện lợi, dễ dàng. Vì yếu tố chức năng ảnh hưởng trực tiếp đến trải nghiệm khách hàng nên cần cải thiện thành phần *tiếp tân* và *dịch vụ phòng*, tăng khả năng tương tác giữa người và robot.
- Yếu tố cảm xúc xã hội dùng để chỉ độ chấp thuận khách hàng với robot. Trong lĩnh vực khách sạn, yếu tố này bao gồm *hiệu ứng nhận diện* và *hiệu ứng cảm xúc*. *Hiệu ứng nhận diện* là hiệu ứng tạo ra bởi sự xuất hiện của robot trong khách sạn, và ảnh hưởng lớn đến trẻ em. *Hiệu ứng cảm xúc* là cảm xúc tạo ra từ cuộc trò chuyện với robot. Khách hàng dễ dàng phân biệt được robot có cảm xúc và không có cảm xúc, tuy nhiên hầu hết robot dịch vụ không đi kèm tính năng này.

- Yếu tố liên quan chỉ độ tin tưởng và độ gần gũi của robot. Trong lĩnh vực khách sạn, yếu tố liên quan chỉ bao gồm *khả năng tương tác* của robot. *Khả năng tương tác* bị ảnh hưởng bởi các yếu tố: trò chuyện đa ngôn ngữ, biểu cảm gương mặt, biểu thị hành động, nhận diện giọng nói, nhận diện cử chỉ. Tuy chỉ chiếm phần ít trong các bài đánh giá, đây lại là yếu tố mang lại trải nghiệm tiêu cực cho khách hàng cao, chủ yếu đến từ khả năng sử dụng ngôn ngữ của robot.

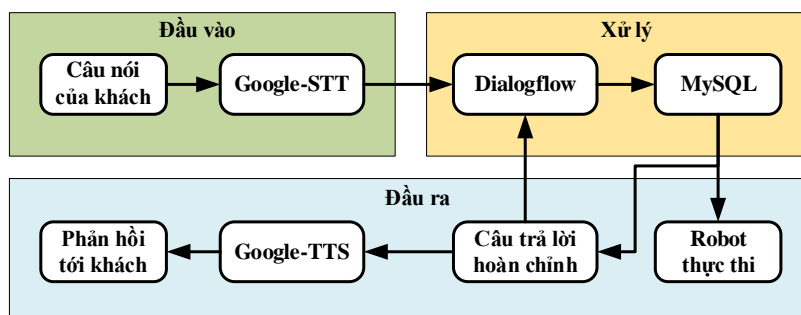
Việc hiểu rõ cơ sở lý thuyết sRAM ứng dụng trong lĩnh vực dịch vụ khách sạn giúp nắm bắt được những hạn chế hiện có của robot dịch vụ hoạt động trong lĩnh vực này. Qua đó, xây dựng một mô hình trợ lý ảo giúp robot tương tác với khách hàng linh hoạt hơn, có mục đích hơn và giải quyết được phần nào các hạn chế đã trình bày.

### B. Hướng tiếp cận

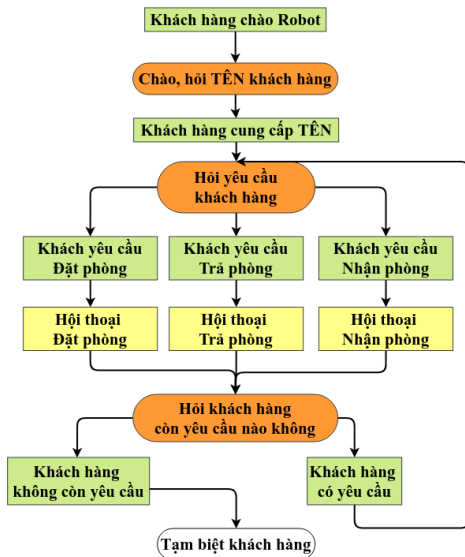
Trợ lý ảo là phần mềm tự động hóa giao tiếp với người dùng và đang ngày càng trở nên phổ biến trong nhiều lĩnh vực. Hiện nay có nhiều framework hỗ trợ xây dựng trợ lý ảo được phát triển bởi các công ty lớn như Dialogflow bởi Google [12], Watson bởi IBM [18], Lex bởi Amazon [19] hay là LUIS bởi Microsoft [20]. Những framework khác nhau có các đặc tính khác nhau. Việc lựa chọn và sử dụng framework hợp lý sẽ giúp tối ưu khả năng trò chuyện cũng như hạn chế các lỗi có thể phát sinh.

Trợ lý ảo cần có khả năng giao tiếp qua giọng nói, với tốc độ phản hồi nhanh và khả năng bảo mật tốt. Dựa trên kết quả so sánh từ [21], Dialogflow, Watson và LUIS là những framework cung cấp cả ba yếu tố trên. Các framework còn lại thiếu khả năng phiên dịch giọng nói (ChatterBot [22], Chatfuel [23]), bị giới hạn ngôn ngữ chỉ gồm Tiếng Anh (Lex [19], Xenioo [24]), khó tích hợp vào các nền tảng khác (Bot Framework [25]), cần chuyên môn cao để phát triển (Knowrob [26]). Đối với ngôn ngữ Tiếng Việt, hiện tại chỉ có Dialogflow hỗ trợ đầy đủ chức năng nhận diện giọng nói, chuyển giọng nói thành văn bản và tách ý nghĩa của câu. Vì vậy, DialogFlow là nền tảng tối ưu giúp dễ dàng phát triển trợ lý ảo sử dụng ngôn ngữ Tiếng Việt.

Với những lợi ích nêu trên, nghiên cứu xây dựng lưu đồ hội thoại cho trợ lý ảo trên nền tảng Dialogflow và



Hình 1: Sơ đồ hệ thống trợ lý ảo



Hình 2: Lưu đồ Dialogflow rút gọn

kết hợp sử dụng các Google Cloud API khác cùng với một số phần mềm mã nguồn mở để đáp ứng yêu cầu liên quan trong sRAM. Chi tiết hệ thống sẽ được trình bày trong phần III.

### III. TRỢ LÝ ẢO HỖ TRỢ ROBOT TIẾP TÂN KHÁCH SẠN

Hình 1 mô tả tổng quan hệ thống trợ lý ảo hỗ trợ tương tác giữa robot tiếp tân khách sạn và khách hàng. Câu nói của khách hàng sẽ được đưa tới khối “Đầu vào”, và được chuyển thành văn bản. Văn bản sau đó được chuyển tới khối “Xử lý”. Tại đây, văn bản được xử lý bởi lưu đồ DialogFlow [12] và đưa ra thông tin phản hồi cho câu nói của khách hàng. Thông tin phản hồi được kết hợp với thông tin truy vấn từ MySQL [11] để đưa ra lệnh điều khiển robot và câu trả lời hoàn chỉnh. Khối “Đầu ra” nhận lệnh và đưa ra yêu cầu điều khiển robot tiếp tân. Đồng thời khối “Đầu ra” cũng cập nhật trạng thái cho Dialogflow và chuyển câu trả lời hoàn chỉnh thành giọng nói để phản hồi lại khách hàng. Chi tiết các module trong từng khối được mô tả rõ hơn dưới đây.

#### A. Xử lý giọng nói

Khối "Đầu vào" thu âm giọng nói khách hàng thành tệp âm thanh. Tệp âm thanh sau đó được gửi dữ liệu lên Google Cloud, đi qua module Google Cloud Speech to Text API (Google-STT) [27] chuyển thành dạng văn bản. Google-STT cho phép chuyển văn bản thành giọng nói từ 125 ngôn ngữ bằng trí tuệ nhân tạo. Tuy có nhược điểm là thời gian dịch vụ phản hồi phụ thuộc tốc độ đường truyền, nhưng lại cho kết quả tốt với ngôn ngữ Tiếng Việt kể cả ở trong môi trường có tiếng ồn, bị nhiễu. Kết quả thực nghiệm cho thấy module Google-STT hoạt động với độ ổn định và tính tin cậy cao.

#### B. Cơ sở dữ liệu

Để quản lý dữ liệu khách sạn như lịch sử đặt phòng, tình trạng phòng, chúng tôi thử nghiệm giải pháp trực tuyến Firebase [28] và giải pháp ngoại tuyến MySQL. Cả hai giải pháp đều cung cấp chức năng quản trị cơ sở dữ liệu, khả năng truy vấn và cập nhật thông tin, nhưng mang tính chất khác nhau. Tuy Dialogflow có thể truy cập trực tiếp vào kho dữ liệu Firebase do cùng một nền tảng Google Cloud, cách thức quản lý dữ liệu NoSQL của Firebase mang lại sự khó khăn trong việc đồng bộ hóa, cũng như tốc độ đường truyền cũng ảnh hưởng đến độ trễ hệ thống. MySQL sử dụng cách thức quản lý dữ liệu SQL, dễ sử dụng, dễ đồng bộ với hệ cơ sở dữ liệu có sẵn của khách sạn. Tuy vậy, để MySQL phản hồi với hệ thống trực tuyến Dialogflow, chúng tôi dùng Webhook với server của bên thứ ba, nên độ trễ đường truyền là không thể tránh khỏi.

#### C. Robot thực thi

Do điều kiện dịch bệnh không cho phép kiểm chứng với hệ thống robot thật, để kiểm tra khả năng nhúng của DialogFlow vào hệ thống robot tiếp tân, chúng tôi thử nghiệm mô phỏng điều khiển robot trên phần mềm Gazebo [13]. Là nền tảng mã nguồn mở về mô phỏng, Gazebo giúp người dùng tạo và điều khiển môi trường 3D mà vẫn đảm bảo các yếu tố vật lý trong thực tế. Trong bài toán này, robot nhận nhiệm vụ tương tác với khách hàng mà DialogFlow yêu cầu ví dụ như giao, nhận chìa khóa phòng khi khách hàng cần check-in, check-out.

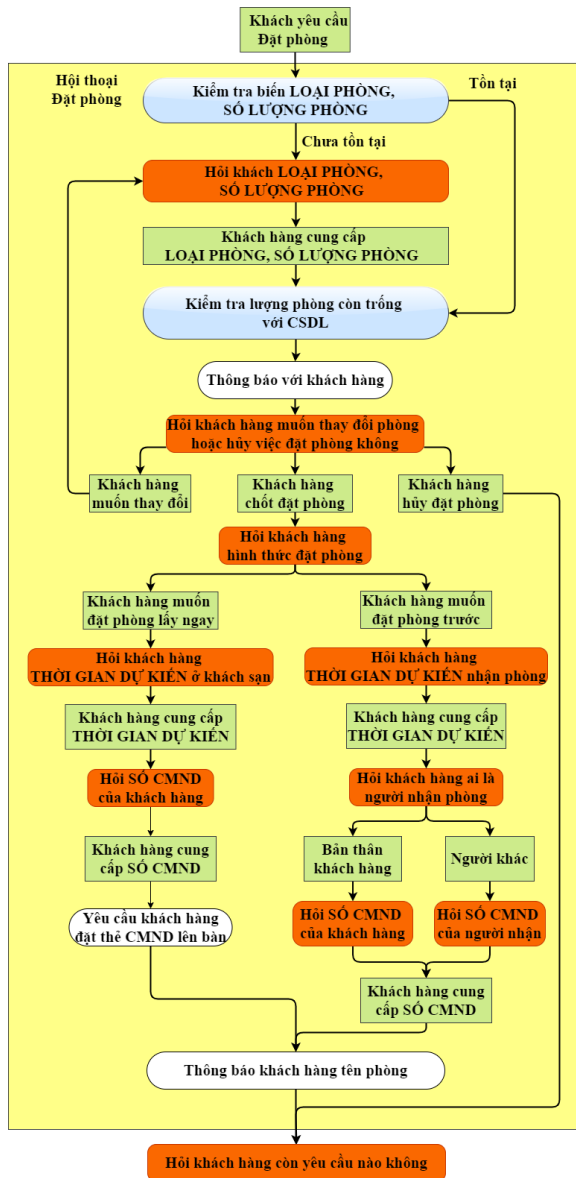
#### D. Phản hồi giọng nói

Sau khi có câu trả lời hoàn chỉnh từ việc kết hợp Dialogflow và MySQL, hệ thống chuyển văn bản thành giọng nói để phản hồi khách hàng. Chúng tôi thử nghiệm việc chuyển văn bản thành giọng nói trên 2 giải pháp trực tuyến là Google Cloud API Text to Speech (Google-TTS) [29] và FPT.ai [30]. Cả hai đều cho độ phản hồi tốt, chất giọng trong trẻo. Tuy nhiên, để đồng bộ với module chuyên đổi giọng nói thành văn bản Google-STT đã trình bày ở trên, nghiên cứu sẽ sử dụng module Google-TTS cho bài toán đặt ra.

#### E. Lưu đồ hội thoại trong Dialogflow

Để đáp ứng yêu tố chức năng và cảm xúc xã hội yêu cầu bởi sRAM cho robot tiếp tân, nghiên cứu thiết kế lưu đồ hội thoại trong Dialogflow như Hình 2. Do kích thước hình ảnh lớn, nên Hình 2 mô tả lưu đồ rút gọn. Lưu đồ hội thoại đầy đủ có thể xem tại: [Lưu đồ hội thoại Dialogflow](#).

Lưu đồ hội thoại thực chất là một đồ thị tạo nên bởi các intent node và function node. Intent node được thể hiện bởi các khối màu xanh lá, nhận đầu vào văn bản từ người dùng, phân tách đối tượng trong câu để xác định mục đích của đầu vào. Function node được thể hiện bởi các khối màu cam, trắng và xanh dương tương ứng với



Hình 3: Lưu đồ Dialogflow chức năng đặt phòng

các chức năng khác nhau. Khối màu cam đưa ra chỉ dẫn và yêu cầu người dùng nhập đầu vào. Khối màu trắng đưa ra phản hồi đến người dùng. Khối màu xanh dương (nằm trong các khối chức năng mô tả ở sau) kiểm tra mối quan hệ của các biến trong hệ thống và CSDL.

Lưu đồ hội thoại bắt đầu sau khi khách hàng chào robot. Sau khi function node yêu cầu khách hàng cung cấp tên và yêu cầu, intent node xác định mục đích khách hàng và di chuyển tới lưu đồ chức năng tương ứng. Sau khi hoàn thành lưu đồ chức năng, trợ lý ảo sẽ hỏi yêu cầu khách hàng hoặc kết thúc trò chuyện nếu khách hàng không còn yêu cầu nào khác.

Các khối màu vàng trên Hình 2 là đại diện cho các lưu đồ hội thoại chức năng với các mục đích khác nhau. Sau đây là giải thích chi tiết hoạt động của các lưu đồ hội thoại chức năng dựa trên hai yếu tố chức năng và cảm xúc.

### 1. Yếu tố chức năng

Dịch vụ tiếp tân bao gồm hai chức năng cơ bản là check-in và check-out. Check-in là hành động người dùng lấy phòng, còn check-out là hành động thanh toán tiền phòng. Sau khi phân tích các mô hình khách sạn thực tế tại Việt Nam, cũng như tham khảo các nhận xét ở [17], chúng tôi chia chức năng check-in thành hai khối “Đặt phòng”, “Nhận phòng”, check-out là khối “Trả phòng”. Trong khối “Đặt phòng”, khách hàng có thể đặt lấy ngay hoặc đặt trước. Khách hàng cần cung cấp loại phòng, số lượng phòng, số CMND và thời gian dự kiến (ở lại/ đến nhận phòng). Khách hàng cũng có thể thay đổi các thông tin trên hoặc hủy việc đặt phòng. Với các khách hàng đặt trước, khi người nhận phòng check-in, khối “Nhận phòng” yêu cầu cung cấp số CMND, tên phòng để đối chiếu với CSDL. Khi khách hàng check-out, khối “Trả phòng” chỉ yêu cầu tên phòng của khách, thực hiện tham chiếu với CSDL để tính tiền phòng. Việc nhận biết các hành động cụ thể giúp lưu đồ trò chuyện đáp ứng đầy đủ các yêu cầu của khách hàng. Trong các trường hợp đặc biệt, lưu đồ trò chuyện có thể mở rộng bằng cách thêm bớt các hành động. Vì vậy, lưu đồ trò chuyện hoàn toàn đáp ứng yếu tố chức năng của dịch vụ tiếp tân.

Chi tiết lưu đồ chức năng cho hành động “Nhận phòng” được thể hiện ở Hình 3. Tại intent node “Khách yêu cầu nhận phòng”, nếu đầu vào nhận được đã bao gồm phòng khách cần nhận (ví dụ “Tôi muốn nhận phòng 203”) thì lưu đồ sẽ di chuyển theo hướng “Tồn tại” đến function node “Hỏi số CMND của khách hàng”. Nếu không, lưu đồ yêu cầu cung cấp thông tin phòng tại function node và khách hàng trả lời tại intent node theo hướng di chuyển “Chưa tồn tại”. Sau khi khách hàng cung cấp số CMND, hệ thống tiến hành kiểm tra thông tin đã ghi nhận trên hệ thống bao gồm số CMND, tên phòng với thông tin khách hàng cung cấp tại function node “Đối chiếu thông tin với CSDL”. Nếu “Không trùng khớp”, hệ thống thông báo với khách hàng và kết thúc lưu đồ chức năng. Ngược lại, hệ thống di chuyển theo hướng “Trùng khớp”, hỏi khách hàng thời gian dự kiến ở lại khách sạn và yêu cầu đưa thẻ CMND cho robot để hoàn tất nhận phòng. Lưu đồ chức năng hoạt động dựa trên việc kiểm soát thông tin và đưa ra quyết định phù hợp. Intent node là đầu vào của hệ thống, tách thông tin cần thiết từ người dùng. Function node không chỉ là đầu ra hệ thống, hướng dẫn (khối cam), cung cấp thông tin (khối trắng) cho người dùng mà còn hỗ trợ kiểm tra các điều kiện thông tin giữa người dùng và CSDL (khối xanh dương). Nhờ vậy, lưu đồ chức năng luôn đảm bảo khách hàng cung cấp đầy đủ, chính xác thông tin cho các mục đích tương ứng. Lưu đồ cũng cung cấp khả năng tùy chỉnh cao, nhờ việc tùy biến, thay đổi, thêm bớt các node và tạo đường dẫn dễ dàng.

Chi tiết các lưu đồ hội thoại chức năng còn lại được trình bày tại: [Lưu đồ hội thoại Dialogflow](#). Nguyên lý

hoạt động của các lưu đồ còn lại tương tự với lưu đồ “Đặt phòng” đã được trình bày ở trên.

## 2. Yếu tố cảm xúc

Yếu tố cảm xúc xã hội của trợ lý ảo được thể hiện qua khả năng tạo ra “hiệu ứng cảm xúc” với khách hàng. Hệ thống hiện tại có thể mô tả cảm xúc của trợ lý ảo. Ví dụ, sau khi khách hàng đặt phòng, trả phòng thành công, tại các function node màu trắng có thể phản hồi “Tôi rất vui vì đã có thể giúp đỡ bạn”. Hoặc khi khách hàng hủy đặt phòng, đưa ra thông tin không trùng CSDL, trợ lý ảo sẽ phản hồi “Tôi rất tiếc vì sự bất tiện này.”. Dialogflow cũng hỗ trợ đặt nhiều mẫu phản hồi tại một function node, nên trợ lý ảo có thể tạo ra trải nghiệm cảm xúc đa dạng với người dùng.

## IV. KIỂM THỬ VÀ ĐÁNH GIÁ HỆ THỐNG

### A. Kiểm tra độ trễ của hệ thống

#### 1. Google Cloud API

Thử nghiệm đo độ trễ của việc gọi Google Cloud API, từ lúc bắt đầu gửi dữ liệu đến Cloud Server khi nhận được phản hồi. Để đảm bảo điều kiện đường truyền là khách quan với mọi lần đo, chúng tôi sử dụng nền tảng Google Colab [32]. Đây là nền tảng cho phép chạy lệnh ngay trên trình duyệt, sử dụng điện toán đám mây. Chúng tôi thực hiện đo độ trễ của việc gọi Google-STT với độ dài của tệp âm thanh đầu vào thay đổi từ 0.5s-5s và đo độ trễ của việc gọi Google-TTS với số lượng từ trong câu thay đổi từ 1-50 từ. Sau khi chạy thí nghiệm 100 lần, giá trị trung bình độ trễ được thể hiện ở Bảng I.

Kết quả ở Bảng I cho thấy việc gọi Google Cloud API có độ trễ thấp, thích hợp trong ứng dụng tương tác thực tế. Việc ứng dụng Google Cloud API giúp giảm tải

BẢNG I: ĐỘ TRỄ TRUNG BÌNH CỦA API

Độ dài tệp âm thanh (s)	Độ trễ Google-STT (s)	Số từ	Độ trễ Google-TTS (s)
0.5-1.0	0.2-0.6	1-10	0.12-0.16
1.0-2.0	0.5-1.2	11-20	0.14-0.19
2.0-3.0	0.9-1.9	21-30	0.17-0.33
3.0-5.0	1.5-3.6	31-50	0.22-0.69

BẢNG II KẾT QUẢ THỰC NGHIỆM

	Trước bổ sung tập học trước	Sau bổ sung tập học trước
Node có 10 câu	60%-70%	85-100%
Node có 20 câu	52%-70%	80%-95%
Node có 30 câu	65%-80%	85%-93%
Kịch bản “Đặt phòng”	71.1%	90.9%
Kịch bản “Nhận phòng”	74.2%	95.5%
Kịch bản “Trả phòng”	78.1%	85.4%

khối lượng lập trình, dễ dàng thiết lập ngôn ngữ theo môi trường và đạt hiệu quả cao.

## 2. Dialogflow

Tương tự thử nghiệm với Google Cloud API, chúng tôi đo độ trễ từ lúc gửi đầu vào văn bản đến khi nhận được phản hồi văn bản trong các intent node của Dialogflow. Chúng tôi nhận thấy độ trễ tuy phụ thuộc vào độ phức tạp trong việc tách ý nghĩa trong câu, nhưng độ lệch không đáng kể, và độ trễ trung bình nằm trong khoảng 0.1s-0.3s.

Với kết quả này, có thể kết luận Dialogflow API không gây ra độ trễ cho toàn bộ hệ thống. Vì vậy sử dụng Dialogflow API đáp ứng nhu cầu về tốc độ phản hồi của khách hàng trong thực tế.

### B. Kiểm tra độ chính xác của lưu đồ hội thoại

Để kiểm tra độ chính xác của Dialogflow với nhiều tình huống hội thoại khác nhau, chúng tôi kiểm tra độ chính xác của mỗi intent node, và kiểm tra khả năng hoàn thành hội thoại của Dialogflow với kịch bản được sinh ngẫu nhiên. Mỗi intent node xác định đối tượng trong câu dựa trên việc học trước những câu có sẵn. Vì thế, với mỗi intent node, tương ứng với cấu trúc đầu vào được mong đợi, chúng tôi biên soạn tập những câu cùng cấu trúc nhưng thay thế các số từ, đại từ, tên riêng có trong câu, thêm thán từ để điều chỉnh sắc thái của câu (lịch sự, vắn tắt), thêm các từ đồng nghĩa. Ví dụ với đầu vào mong muốn là “Tên tôi là A”, chúng tôi thay thế “tôi” bằng “chị”, “anh”, “mình”; thay thế “A” bằng các tên riêng khác; thêm “ạ”, “vâng”, “nhé”. Các kịch bản hội thoại tạo ra qua hai bước như sau:

**Bước 1:** Chọn hành động cần kiểm tra, từ đó chọn ra danh sách các intent node.

**Bước 2:** Với mỗi intent node, chọn ngẫu nhiên 1 câu trong tập câu cùng cấu trúc.

Mỗi intent node, được kiểm tra với 10-30 câu tương tự cấu trúc. Mỗi hành động được kiểm tra với 1000 kịch bản. Sau khi chạy chương trình trên Google Colab, chúng tôi nhận thấy độ chính xác có thể cải thiện bằng cách gia tăng tập học trước của mỗi intent node. Kết quả độ chính xác được thể hiện ở Bảng II.

Kết quả ở Bảng II chỉ ra rằng lưu đồ hội thoại có độ chính xác cao, đáp ứng được nhiều tình huống tương tác trong thực tế. Lưu đồ hội thoại cũng có thể điều chỉnh theo mong muốn của khách sạn, cũng như tiếp tục cải thiện độ chính xác bằng cách gia tăng số lượng và chất lượng của tập học trước.

### C. Kiểm tra quy trình tổng thể của trợ lý ảo

Kết quả kiểm tra tổng thể quy trình hoạt động của trợ lý ảo được ghi hình và trình bày tại: [Kết quả thực nghiệm](#)

Thông qua video kiểm thử hoạt động, trợ lý ảo có thể đáp ứng tốt các yêu cầu Đặt phòng lấy ngay, nhận phòng và trả phòng với thời gian trễ ngắn, độ chính xác cao. Trợ lý ảo cũng có thể kết nối với robot mô phỏng để cung cấp thẻ phòng hoặc thu thẻ CMND. Vì vậy

hoàn toàn có thể áp dụng trợ lý ảo vào mô hình thực tế và nâng cao trải nghiệm khách hàng.

## V. KẾT LUẬN

Bài báo trình bày hệ thống trợ lý ảo hỗ trợ robot tiếp tân khách sạn trong việc tương tác với khách hàng. Kết quả kiểm thử cho thấy hệ thống có độ trễ thấp, độ chính xác cao và có thể cải thiện theo thời gian. Nhờ việc áp dụng cơ sở lý thuyết sRAM, trợ lý ảo được trình bày có thể khắc phục một số yếu điểm như chưa có cảm xúc và độ tương thích, khả năng đáp ứng chưa cao của một số hệ thống hiện tại. Theo đó nâng cao chất lượng phục vụ tại bàn tiếp tân và trải nghiệm của khách hàng.

Tuy nhiên, trợ lý ảo mới khai thác các chức năng của Dialogflow đang dừng lại ở mức căn bản, cũng như chưa đạt độ chính xác tuyệt đối. Trong tương lai, chúng tôi sẽ tiếp tục hoàn thiện lưu đồ hội thoại, thêm chức năng tự động học trước cũng như khảo sát độ tin cậy của trợ lý ảo trên môi trường thực tế.

## LỜI CẢM ƠN

Đình Bảo Minh được tài trợ bởi Công ty cổ phần VinBigdata thuộc Tập đoàn Vingroup và hỗ trợ bởi chương trình học bổng đào tạo thạc sĩ, tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), Viện Nghiên cứu Dữ liệu lớn (VinBigdata), mã số VINIF.2020.ThS.43.

## TÀI LIỆU THAM KHẢO

- [1] L. Zhong, S. Sun, R. Law and X. Zhang, "Impact of robot hotel service on consumers' purchase intention: a control experiment," *Asia Pacific Journal of Tourism Research*, vol. 25, no. 7, pp. 780-798, 2020.
- [2] J. Hwang and S. Seo, "A critical review of research on customer experience management: Theoretical, methodological and cultural perspectives," *International Journal of Contemporary Hospitality Management*, vol. 28, 2016.
- [3] E. ZalamaEmail, J. G. García-Bermejo, S. Marcos, S. Domínguez, R. Feliz, R. Pinillos and J. López, "Sacarino, a Service Robot in a Hotel Environment," *ROBOT2013: First Iberian Robotics Conference*, vol. 253, 2014.
- [4] S. Ivanov, C. Webster and A. Garenko, "Young Russian adults' attitudes towards the potential use of robots in hotels," *Technology in Society*, vol. 55, pp. 24-32, 2018.
- [5] D. Belanche, L. V. Casalo, C. Flavián, J. Schepers, "Service robot implementation: a theoretical framework and research agenda," *The Service Industries Journal*, vol. 40, pp. 203-255, 2020.
- [6] S. Mahmud, R. H. Sourave, M. Islam, X. Lin and J.-H. Kim, "A Vision based Voice Controlled Indoor Assistant Robot for Visually Impaired People", in 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1-6, Sept. 2020.
- [7] H. S. Ahn, W. Yep, J. Lim, B. K. Ahn, D. L. Johanson, E. J. Hwang, M. H. Lee, E. Broadbent and B. A. MacDonald, "Hospital Receptionist Robot v2: Design for Enhancing Verbal Interaction with Social Skills", in 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Oct. 2019.
- [8] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch and A. Martins, "Brave new world: service robots in the frontline," *Journal of Service Management*, vol. 29, no. 5, 2018.
- [9] V. W. S. Tung and R. Law, "The potential for tourism and hospitality experience research in human-robot interactions," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 10, 2017.
- [10] Google, "Google Cloud," Google, [Online]. Available: <https://cloud.google.com/>.
- [11] MySQL, "MySQL," MySQL, [Online]. Available: <https://www.mysql.com/>.
- [12] Google, "Dialogflow," Google Cloud, [Online]. Available: <https://dialogflow.cloud.google.com/>.
- [13] Gazebo, "Gazebo," ©2014 Open Source Robotics Foundation, [Online]. Available: <http://gazebo.org/>.
- [14] V. Viswanath, M. G. Morris, G. B. Davis and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View," vol. 27, no. 3, 2003.
- [15] H. Aoki, Y. Fujimoto, S. Suzuki, E. Shimokawara and T. Yamaguchi, "Difference in physiological responses by different cultural greetings using a robot," In *International Workshop on Advanced Computational Intelligence and Intelligent Informatics (Suzhou, China)*, 2011.
- [16] D. Li, P. L. P. Rau and Y. Li, "A Cross-cultural Study: Effect of Robot Appearance and Task," *International Journal of Social Robotics*, vol. 2, no. 2, 2010.
- [17] L. Fuentes-Moraleda, P. Díaz-Pérez, A. Orea-Giner, A. M.-Mazón and T. Villacé-Molinero, "Interaction between hotel service robots and humans: A hotel-specific Service Robot Acceptance Model (sRAM)," *Tourism Management Perspectives*, vol. 36, 2020.
- [18] IBM, "IBM Watson," IBM, [Online]. Available: <https://www.ibm.com/watson>.
- [19] Amazon, "Amazon Lex," Amazon, [Online]. Available: <https://aws.amazon.com/vi/lex/>.
- [20] Microsoft, "Microsoft Azure," Microsoft, [Online]. Available: <https://www.luis.ai/>.
- [21] S. Perez-Soler, S. Juarez-Puerta, E. Guerra and J. d. Lara, "Choosing a Chatbot Development Tool," *IEEE Software*, vol. 38, no. 4, pp. 94-103, 2021.
- [22] ChatterBot, "ChatterBot," ChatterBot, [Online]. Available: <https://chatterbot.readthedocs.io/>.
- [23] Chatfuel, "Chatfuel," Chatfuel, [Online]. Available: <https://chatfuel.com/>.
- [24] Xenioo, "Xenioo," Xenioo, [Online]. Available: <https://www.xenioo.com/>.
- [25] Microsoft, "Microsoft Bot Framework," Microsoft, [Online]. Available: <https://dev.botframework.com/>.
- [26] Knowrob, "Knowrob," Knowrob, [Online]. Available: <http://knowrob.org/>.
- [27] Google, "Cloud Speech-to-Text," Google, [Online]. Available: <https://cloud.google.com/speech-to-text>.
- [28] Google, "Firebase," Google, [Online]. Available: <https://firebase.google.com/>.
- [29] Google, "Cloud Text-to-Speech," Google, [Online]. Available: <https://cloud.google.com/text-to-speech/>.
- [30] FPT, "FPT.ai," FPT, [Online]. Available: <https://fpt.ai/>.